A BAYESIAN APPROACH TO ASSIST IN THE DIAGNOSIS OF CORONARY HEART DISEASE

William Randolph Condos

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A BAYESIAN APPROACH TO ASSIST IN
THE DIAGNOSIS OF CORONARY HEART DISEASE

by

William Randolph Condos, Jr.

and

Everett William Knox

Thesis Advisor: M. U. Thomas

March 1973

Approved for public release; distribution unlimited.

A Bayesian Approach to Assist in the Diagnosis of Coronary Heart Disease

by

William Randolph Condos, Jr.
Captain, United States Army
B.S., United States Military Academy, 1967

Everett William Knox
Captain, United States Army
B.E.E., Clarkson College of Technology, 1963

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1973

ABSTRACT

The objectives of this thesis were to design a method for evaluation of the diagnostic potential of available indicators of coronary heart disease (CHD) and to present a systematic, quantitative procedure for aiding in its diagnosis. A sample space of patients was divided into two mutually exclusive groups, those with angiographic evidence of CHD, and those with no CHD. Active duty or retired military men between the ages of 30 and 67 years constituted the sample space. Tests and risk factors were available in the medical literature that a doctor could view as an indicator or contraindicator of CHD. A vector of these possible indicators was established and the diseased group was compared to the non-diseased group in an effort to evaluate the diagnostic potential of the indicators. This was don by discriminant analysis in conjunction with a Bayesian method of weighting the importance of test results. The important indicators were then used to formulate a model for diagnosing CHD based on a Bayes' decision technique.

TABLE OF CONTENTS

I. INTRODUCTION

Heart attacks resulting from coronary heart disease (CHD) cause more deaths each year than cancer, strokes, and accidents combined. These deaths also include a broader spectrum of the population than in previous years. In the last century, heart disease was viewed as a natural result of growing old. But with the transition from a rural to an urban society, and the inherent traits of tension, rich diet, and lack of exercise, the propensity for heart disease has increased. This increase can be seen in the steady rise in the number of heart attacks among men over the past 20 years. The American Heart Association reported that of the 675,000 deaths from CHD expected during the past year, 176,000 would have been men and women under the age of 65 [Ref. 16].

Medical capabilities have greatly increased, giving coronary heart disease patients a greater probability of survival once they are under medical care, but since over half of those who die never reach a hospital, the problem of predicting coronary heart disease becomes very important. This diagnostic problem gains additional importance because of the lack of a proven method for the treatment of CHD in its advanced stages. Furthermore, there is an increased presence of asymptomatic CHD that may go undetected with present diagnostic criteria.

In this study an attempt has been made to consolidate a spectrum of risk factors that can be incorporated into diagnostic procedures for CHD. Specifically, the objectives were to design a method for evaluation of the diagnostic potential of available indicators of CHD and to present a systematic, quantitative procedure for aiding in its diagnosis.

A sample space of patients was divided into two mutually exclusive groups, those with angiographic evidence of CHD, and those with no CHD. Active duty or retired military men between the ages of 30 and 67 years constituted the sample space. There were certain tests and risk factors available in the medical literature that a doctor could view as an indicator or contraindicator of the disease. Having established a vector of these possible indicators, the diseased group was compared to the nondiseased group in an effort to evaluate the diagnostic potential of the indicators. This was done by discriminant analysis in conjunction with a Bayesian method of weighting the importance of test results. The important indicators were then used to formulate a model for diagnosing CHD based on a Bayes' decision technique.

II. BACKGROUND

Probabilistic and computer aided designs to aid decision makers in medical diagnosis have been a promising area of research for some time, and an abundant literature on these subjects exists [Refs. 8, 10]. They have had little impact on the practice of medicine, however, with several characteristic reasons being given. Among them may be mentioned insufficient data bases because of the poor quality, lack of uniformity, or inaccessability of medical records. In addition, there appears to be a lack of understanding and interface between the medical profession and those who would apply probabilistic procedures to aid the medical decision makers.

Recent years have shown an increase in research efforts aimed at the prevention and diagnosis of CHD. At the present time, however, coronary arteriography appears to be the only completely definitive test for the disease [Refs. 4, 12]. Unfortunately, this is a costly surgical procedure that requires hospitalization and involves definite mortality and morbidity factors, depending on the age and health of the patient. Arteriography is currently only available at large medical centers because of the equipment and expertise required.

Some diagnostic models for CHD tend to consider only symptomatic patients, usually those with typical angina.

This omits many subjects who are asymptomatic, a portion of which may be suffering from silent heart disease.

The medical literature cites commonly accepted indicators for CHD. Widely used indicators cited are history of ischemic episodes, age, total cholesterol, triglycerides, resting EKG, smoking, and family history [Refs. 4, 12, 16]. Less commonly used indicators that are also cited are race, blood type, and blood pressure [Refs. 4, 9]. In addition, the exercise test has recently gained widespread acceptance as a good CHD indicator [Refs. 1, 6]. The relative importance of this test in conjunction with other indicators has not yet been thoroughly investigated.

It seems appropriate that a diagnostic model for predicting CHD should investigate the potential of an exhaustive list of indicators and tests for the disease. This diagnostic model should also reduce the subjectivity in the decision making of the doctor by increasing the amount of objective evidence through the appropriate indicators and tests.

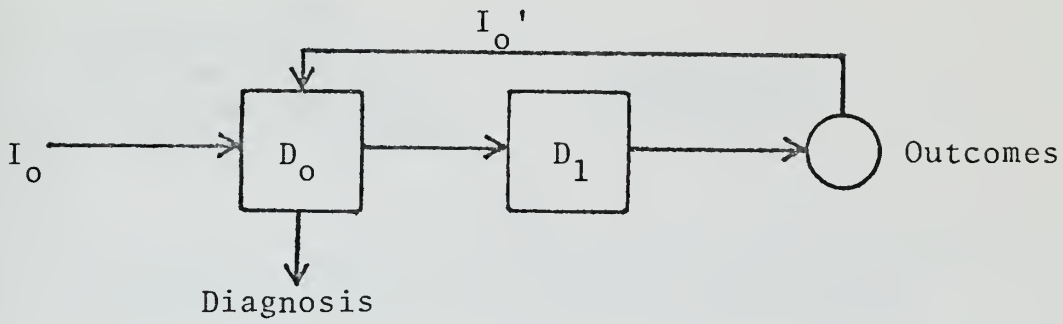
III. DESCRIPTIVE MODEL

The flow of patients to a cardiac clinic is similar to the input of any other specialty clinic. A patient may be referred to the cardiologist by another doctor based on the results of a physical examination or, if a person believes that he is suffering from a cardiac or cardiac-related illness, he may voluntarily seek the advice of the specialist directly. In either case, by the time a patient is admitted to the cardiologist's office, there is already certain data on him that is available to the physician without specified testing. From that point on, however, the diagnosis of a possible heart disease is a function of the doctor's ability to assign relative importance to the appropriate indicators. Costs of associated testing, the procedures available, the patient, and the patient's health may also have a bearing on the doctor's ability to diagnose correctly.

The cardiologist then may be viewed as a decision maker who, for each patient, receives an amount of initial information I_o from which he initiates a sequence of decisions, gaining additional information I_o' as a result of testing. Figure 1 shows a schematic of these decision processes.

FIGURE 1

DECISION PROCESSES OF CARDIOLOGIST



As an illustration of the concepts implied in Figure 1, consider that a patient is referred to the cardiologist because he has symptoms of CHD. At decision node D_o the doctor evaluates the information he has available. Usually this is information readily available in the patient's medical record. Based on this information, the doctor has two choices at D_o, diagnosis of the patient or requesting additional testing. If, for example, the doctor chooses t perform a test, decision node D_1 represents the choice the doctor must make from the clinical tests available. Having made the choice, I_o' represents the information that results from the outcome of the test. The doctor is again faced with the decision to be made at D_o, but he now has the new information I_o' which reduces the chance of an incorrect diagnosis.

A summary is presented in Table 1 that shows the possible path of a patient through a diagnostic sequence.

TABLE 1

PATIENT ADMITTED TO THE CARDIOLOGIST

Race, Sex, Age, Height, Weight, Blood Pressure, Blood Type, Family History of Heart Disease, Smoking History, History of Ischemic Episodes	AVAILABLE INFORMATION (I_o)	(A)

FURTHER TESTING SPECIFIED BY CARDIOLOGIST

Resting EKG Exercise EKG Triglycerides Cholesterol Angiogram	CLINICAL TESTS (I_o')	(B)

This summary does not dictate a specified sequence of tests or weightings of relative importance. The information in (A) is data available (facts about the patient) that are easily obtained without testing. Tests in (B) require expert judgment or clinical procedures and, again, are not ordered in any sequence of importance. In practice, not all of the listed indicators are used for decision making. Some may be considered by a particular doctor to be unimportant. It is also difficult to assign subjective probabilities to some of the indicators about which little is known. Furthermore, it is impractical to correlate the contributions of a large number of indicators without some type of objective model.

IV. QUANTITATIVE METHODS

In general, there are two approaches to medical decision problems. The first is to develop and perfect a model that predicts as well as or better than a physician. The second approach consists of improving ways to aggregate, weight, and use information available to the physician so that his personal diagnosis will be conducted from a substantially sounder base. This latter approach, which is commonly called "bootstrapping" [Ref. 10] was the one selected for this study.

A set of CHD indicators was identified and evaluated experimentally using discriminant analysis. A proposed method of assigning weighting factors based on the "posterior odds" of the various indicator levels was incorporated into the analysis. These results were then integrated into a Bayesian diagnostic model.

A. INDICATORS AND WEIGHTING FACTORS

At decision node D_1 of Figure 1, the doctor must decide what test to use next in his evaluation of the patient. To do this he must have a knowledge of what indicators of CHD have been evaluated and the amount of additional information, I_o', he can expect to obtain from these indicators. Complicating the doctor's evaluation is the division of the indicators into two types, qualitative and quantitative. The quantitative indicators are tests in which the outcome

is represented on an acceptable numerical scale. Of the indicators used in this paper, only triglycerides, cholesterol, age, and blood pressure were quantitative variables. The other indicators shown in Table 1 (except height and weight which were not used) have results which have no numerical scale and must be interpreted qualitatively. For example, the indicator called history of ischemic episodes requires the patient to verbalize his history of chest pain. Also included in the category of qualitative indicators are tests in which the result is numerical but lacks meaning unless expressed in qualitative terms. The exercise EKG result, for example, is in millimeters of depression (or elevation) of the S-T segment, but is interpreted in terms of being positive or negative.

As pointed out previously, these indicators and their relative merit were determined from clinical judgment and varied among cardiologists. In addition, the relative importance of various outcomes of any specific test also varied among doctors. To alleviate these problems, a two-step procedure was used. First, the outcomes of the qualitative tests were assigned weighting factors using Bayes' Theorem. Second, the qualitative variables and quantitative variables were integrated into a relative ranking using a stepwise discriminant analysis computer routine [Ref. 11].

Consider a particular qualitative variable i for which $P(t_{ij}|\bar{D})$ is the conditional probability of outcome j

given a patient has CHD. The posterior probability of CHD (i.e., in light of this information) is

$$P(D|t_{ij}) = \frac{P(t_{ij}|D)P(D)}{P(t_{ij}|D)P(D) + P(t_{ij}|\bar{D})P(\bar{D})} \qquad (1)$$

where P(D) is the presumably known prior probability of CHD. Each of these probabilities on the right hand side of equation (1) can be estimated from past data. The results are a vector of values for the outcomes of a specific test which could then be used with the outcomes of other tests in a stepwise discriminant analysis computer routine. However, in order to give more meaning to the weighting factors, w_{ij}, they were normalized using

$$w_{ij} = \frac{P(D|t_{ij})}{\min_k \{P(D|t_{ik})\}} \qquad (2$$

where it was arbitrarily decided to use the minimum outcome in order to show increasing likelihood of disease as the value of the weighting factor increased.

Consider the following simple example to illustrate the procedure for computing weighting factors. Suppose it is desirable to find weighting factors for the qualitative variable "race" (i = R) which for the purpose of illustration, has two outcomes: NEGRO (j = 1) and CAUCASIAN (j = 2). Suppose further that the prior distribution of CHD is

$P(D) = 0.1$ and data reveals that $P(t_{R1}|D) = 0.2$ and $P(t_{R1}|\bar{D}) = 0.4$. It then follows from equations (1) and (2) that the weighting factors are $w_{R1} = 1.0$ and $w_{R2} = 2.46$.

This method of computing the weighting factors $\{w_{ij}: i=1, \ldots,n; j=1,\ldots,m\}$ provides a consistent means of assigning scores to each of the qualitative variables. This was done for a particular set of indicators examined in this study and the results are given in Section VI. Stepwise linear discriminant analysis [Ref. 11] could, at this point, be used to develop a linear prediction function $L = \sum_{i=1}^{m} \lambda_i X_i$ where X is the set of all test variables (quantitative and qualitative), λ is the set of all coefficients assigned by the computer routine, and m is the number of tests. Mahalanobis distance could then be used as the discrimination criterion.

Cohn [Ref. 4] used this type of linear discriminant analysis in its predictive role in a medical decision context. Use of discriminant analysis for prediction was discarded in this paper for two reasons. The technique is a valid one when the underlying distributions of the random variables of the two samples (in this case, the test results) are distributed normally with equal covariance matrices (a linearity assumption). A preliminary investigation indicated that the variance of the test results in the two samples did not appear to be equal. Additionally, the normality

assumption did not appear to be valid in this application. The test results had a combination of binomial, multinomial, and approximately normal distributions. Consideration of all distributions as normal did not have a sound theoretical basis.

The actual purpose of conducting this portion of the analysis was to identify the relative importance among the variables. This was accomplished by ordering the resulting F-statistics associated with the coefficients (λ's) of the variables (X's). The F-statistic is the ratio of the variability of the means of the individual test results in each sample to the pooled variance of the test results. F will be large when there is a large difference between the mean results of a test in the CHD and the no CHD groups. Likewise, the smaller the F, the closer together are the mean results for a particular test in the CHD and no CHD groups. Thus, an ordering of these computed F-statistics from largest to smallest may be considered an ordinal ranking of the diagnostic power of the various indicators.

B. BAYESIAN DIAGNOSTIC MODEL

The foregoing procedure, of Section IV.A., for determining the relative diagnostic power of the available tests of indicators provides criteria for the cardiologist to select appropriate tests at decision node D_1 in Figure 1. A Bayesian method for quantifying the information I_o and additional information I_o' is now presented.

The development of this model was based on two major assumptions. First, it was assumed that patients being tested either had CHD or did not have CHD. Thus, the case of a patient having multiple diseases was excluded here. The second assumption was that the data, on both qualitative and quantitative variables were conditionally independent.

Let

$P(D_i) \equiv$ apriori probability of CHD (D_1), or no CHD ($D_2=\bar{D}_1$).

$P(D_i|S_1,\ldots,S_n)$ a posterior probability of D_i given symptoms, or indicator levels, $S_1,\ldots,S_n$.

$P(S_1,\ldots,S_n|D_i) \equiv$ conditional probability of symptoms $S_1,\ldots,S_n$ given D_i.

The first assumption merely requires that $P(D_i)=P(D_1)$ or $P(D_2) \equiv P(\bar{D}_1)$. The second assumption, in terms of the above notation, says that

$$P(S_1,\ldots,S_n|D_i) = \prod_{j=1}^{n} P(S_j|D_i) \qquad (3)$$

It then follows from Bayes' Theorem that

$$P(D_i|S_1,\ldots,S_n) = \frac{P(D_i)\prod_{j=1}^{n} P(S_j|D_i)}{\sum_{i=1}^{2} P(S_1,\ldots,S_n|D_i)P(D_i)}$$

which is in terms that can be calculated using subjective probabilities (doctor's medical opinions) and frequentistic procedures [Ref. 8].

The majority of the conditional probabilities were calculated using frequentistic procedures. Subjective probabilities were used when the data base was insufficient. In cases where a patient was missing the j^{th} symptom on his medical records or the patient was unable to take the test, the conditional probabilities $P(S_j|D_1)$ and $P(S_j|D_2)$ were set equal to .5 (i.e., $P(S_j|D_1)$ and $P(S_j|D_2)$ were equally likely and thus had no influence on the associated probabilities).

The Bayesian diagnostic model was developed because it provided several distinct advantages over general discriminant analysis techniques commonly used for medical decision making. The first advantage was the use of subjective apriori probabilities. Each doctor has his own feelings and experience concerning the probability of CHD in a patient. The second advantage was that the Bayesian model is self-updating. After each patient has been diagnosed, his characteristics can be easily added to the data providing new apriori probabilities. This allows the doctor to see trends that may develop, providing the stimulus for research in these areas. The data base is continuously enlarged in this manner, improving the diagnostic accuracy of the model. The third advantage is that CHD is only a small part of the

diagnostic problem facing the doctor. The Bayesian approach allows for the expansion of the hypothesis. In the present model only one hypothesis is treated, no CHD or CHD. However, this could easily be expanded to no disease, CHD, liver disease, etc. An important aspect of this is that as the number of data points in the data vector and the number of hypotheses are increased, the accuracy of the model improves.

V. CLINICAL TESTS AND OBSERVATIONS

Data was derived from three sources. The first source was generated by testing a sample of individuals undergoing routine physical examinations at Fort Ord Army Hospital. A collection sheet was developed to record the data that was simple yet comprehensive enough to see if trends developed in areas not considered important in the initial analysis (See Appendix A).

The second source of data was the medical records at Letterman General Hospital, San Francisco. A data sheet similar to that of the Fort Ord sample was used. However, several problem areas were encountered. The first was the problem of definition and interpretation. Many records showed information such as "positive" family history with no explanation of what the doctor's opinion was based on. Others had entries such as "30 pack year history" of smoking. This type of data does not differentiate between two packs per day for 15 years or three packs per day for 10 years. Since intensity of smoking may be an important variable, much valuable data were lost. Another problem in this area was the omission of data that were assumed to be normal. If a patient's test result was abnormal, the result was noted in the patient's record. (However, if nothing was noted, it was not clear whether the test result was normal or that the result was omitted.) It is clear

that personalities become an important factor in the writing and in the reading of medical records. However, it is felt that as more records are automated these problems will be greatly reduced.

The problem of missing data was the major obstacle encountered from the CHD population. The majority of the patients did not have all the test results in their files. The only solution to this problem is to increase the sample size so that patients with missing data can be removed from the sample. But since one of the major objectives of this paper was to develop a method, the missing data problem will not be considered within this framework. For information concerning decision making with missing data, see Ref. 4.

The third source of data was the medical literature. This was used to establish apriori probabilities of CHD when it was felt that the experimental sample was too small, making the sample probabilities very sensitive to error [Ref. 7].

The partitioning of the sample space into two parts, CHD and no CHD, implied that the subject in the healthy group was not suffering from any disease, and that a subject in the CHD group was suffering from CHD only. Other diseases may have adversely affected the test results of either group. In the formation of the sample, care was taken to eliminate all subjects that had other diseases.

In the determination of positive or negative family history, the age of 65 was considered the cut-off. If a

blood relative had CHD prior to age 65, the result was positive. Although this cut-off was arbitrary, it was the one most consistent with the available literature. It can be easily changed, however, if another cut-off is desired.

When checking for chest pain, the existence of any chest pain that was not categorized as angina was listed as undetermined origin since none of the subjects were known to have diseases which might explain the pain.

The reading of the resting EKG was done by a cardiologist whose experience and subjective opinions must be considered an important part of the data.

IV. SENSITIVITY

Sensitivity analysis was conducted in the following areas:

1. The effect of weighting factors on the ordinal ranking of the qualitative indicators was investigated. Table 2 shows how changes in weighting factors proved to markedly influence the diagnostic ordering of the indicators shown in Table 3.

TABLE 2

Test	Bayes' Weighting Factor	Sample Clinical Judgment Weighting Factor
Blood Type		
A	1.3	2
Other	1	1
Family History		
Positive	1.3	2
Negative	1	1
Smoking History (per day)		
Non-smokers	1	1
Less than 1/2 pack	4.6	2
About 1 pack	4.5	3
Greater than 1 pack	6.3	4
History of Ischemic Episodes		
None	1	1
Chest pain	8	2
Typical angina	335	3

TABLE 2 (Continued)

Resting EKG		
Normal	1	1
Other	4	2
ST-T abnormalities	20	3
Pathologic Q-waves	22.5	4
Race		
Caucasian	6.5	1
Negro	1	2
Mongolian	1.5	3
Exercise EKG		
Normal	1	1
ST depression < 1mm	25	2
ST depression $\geq$ 1mm	150	3

All other indicators were quantitative. The following ordering of indicators and their associated F-statistics resulted (Table 3):

TABLE 3

Bayesian Weighting Procedure	
History of Ischemic Episodes	97.5709
Exercise EKG	5.3225
Age	3.2315
Resting EKG	2.6744

TABLE 3 (Continued)

Blood Type	2.6532
Cholesterol	2.2091
Density	2.0640
Cigarette Smoking	1.8119
Systolic Blood Pressure	.9659
Family History	.1607
Triglycerides	.0485

Diastolic Blood Pressure and Race were omitted because of an insignificant F value for this particular sample.

Sample Clinical Weighting Procedure

Exercise EKG	13.0179
History of Ischemic Episodes	7.0552
Density	2.3373
Blood Type	2.2919
Cholesterol	2.1914
Systolic Blood Pressure	2.1383
Resting EKG	1.7011
Cigarette Smoking	1.5936
Family History	1.1054
Diastolic Blood Pressure	.2963
Triglycerides	.2573
Age	.0243
Race	.0169

2. Diagnostic accuracy was investigated by varying the prior probability of disease, P(D), and assuming $P(D|S_1,\ldots S_n) > 0.5$ indicated CHD. These values for Table 4 were determined from patients having eight or more test results.

TABLE 4

P(D)	False Negatives**	False Positives***
.04	9/50	0/52
.10	6/50	0/52
.20	5/50	0/52
.30	5/50	1/52
.40	3/50	1/52
.50	2/50	1/52

** False Negative ≡ patient has CHD but is diagnosed as not having CHD.

*** False Positive ≡ patient does not have CHD but is diagnosed as having CHD.

3. After the model had been developed and the conditional probabilities had been determined, data on CHD patients were obtained from Walter Reed Hospital. Using the originally determined probabilities, these patients were tested with the Bayes' diagnostic model and 12 out of 14 were correctly diagnosed as having CHD. Again, a $P(D|S_1,\ldots, S_n) > 0.5$ indicated CHD.

The Walter Reed patients were then added to the original sample to update the prior probability of disease. The

changes in the prior probabilities were so small that they had no effect on the diagnostic results.

4. Diagnostic accuracy was investigated by varying that probability above which CHD would be indicated (Table 5):

TABLE 5

$P(D \mid S_1,\ldots,S_n)$	False Negatives	False Positives
.1	3/58	1/52
.2	4/58	0/52
.3	8/58	0/52
.4	9/58	0/52
.5	9/58	0/52
.6	9/58	0/52
.7	9/58	0/52
.8	11/58	0/52
.9	15/58	0/52

VII. RESULTS AND CONCLUSIONS

As previously stated in Section I, the objectives of the study were to design a method for the evaluation of the diagnostic potential of available indicators of CHD and to present a systematic, quantitative procedure for aiding in its diagnosis. The indicators of CHD were investigated by comparing specific test results from a CHD sample and a healthy sample with no CHD.

The stepwise discriminant analysis, as presented in Section IV.A., using all variables was performed on a CHD sample size of 106 compared to a no CHD sample size of 56. The weighting factors were determined by the Bayesian approach (tabulated in Table 3, Section VI). An important result of the discriminant analysis program was the ordering of variables and their associated F-statistics which may be viewed as an ordering of the relative diagnostic importance of the tests (see Table 4, Section VI). This method of assigning weighting factors to test results in conjunction with discriminant analysis is a valid procedure for ordering the vector of tests in their diagnostic importance. It provides a means for a doctor at decision node D_1 (of Figure 1) to determine which test provides the most additional information I_o' from those available to him. Additionally, the method is particularly valuable and easily adapted to considering new indicators of disease where no definitive

clinical judgment exists or doctors do not agree on the relative importance of test results.

The Bayes' diagnostic model (Section IV.B.) was developed to provide a systematic, quantitative procedure for aiding in the diagnosis of CHD. It was evaluated by checking how well it diagnosed patients from a known CHD group and a known healthy group. The difficulty in obtaining patients with all the required test results was noted in Section V and resulted in extremely small samples with complete data to investigate. However, six out of seven of the CHD group were diagnosed correctly, and 33 out of 33 of the no CHD group were diagnosed correctly. When only eight or more of the test results were available, the model diagnosed with 91% accuracy (41 out of 50 in the CHD group were diagnosed correctly and 52 out of 52 of the no CHD group were diagnosed correctly). These results were based on using a posterior probability of disease of .50 as the cut-off probability (i.e., $P(D|S_1,\ldots,S_n) \geq .50$ indicated CHD). The variation of the cut-off probability (see Section VI) demonstrated that the diagnostic accuracy of the model was greatly influenced by the choice of the cut-off criterion. For example, using a cut-off of .20 instead of .50 reduced the number of false negatives from nine to four while the number of false positives remained the same.

As a validation of the Bayes' diagnostic model, 14 known CHD patients from Walter Reed Hospital were diagnosed

by the model. Twelve of the 14 were diagnosed correctly. The validation is not conclusive because of the extremely small sample tested, but it does indicate that the method is promising.

It may be desirable to use the methods presented in a screening program to identify people with high risk of CHD from a large population. Sufficient doctors may not be available to examine all of the people to be tested. As an example of the model's applicability to such a screening program (where a doctor is not required) diagnostic accuracy was investigated using the results of the information available only [referred to in Figure 1 as I_o and in Table 1 as (A)]. The model diagnosed with 92% accuracy (19 out of 24 in the CHD group were diagnosed correctly and 44 out of 44 in the no CHD group were diagnosed correctly).

The Bayesian diagnostic model had a high degree of accuracy in correct diagnoses. It is easily implemented and appears to be well adapted to screening studies where a large population is involved. The model continuously updates the available patient information from which the conditional probabilities are calculated and may be useful in indicating trends or fluctuations in the indicators of disease.

VIII. AREAS FOR FUTURE STUDY

As pointed out previously (see Section IV.A), one of the main advantages of the approach followed in the paper is the easy expansion of the number of variables and the number of patients to be tested. This implies that as the number of variables is increased, the diagnosis of CHD will improve. The expanded list of variables could also be used to predict other diseases. Instead of a space of CHD and no CHD, there is a space of CHD plus other diseases limited only by logical considerations such as the time, money, availability of computational equipment, etc. The integration of this expanded prediction model into routine physical examinations and patient history could allow preliminary diagnosis prior to consultations with doctors, helping to reduce costs and the increasing patient load of doctors.

As presently modeled, diagnosis is based on results of samples from diseased and non-diseased groups. However, as more samples are obtained and a history of the patient's variables (i.e., changes in blood pressure over several years) is made, the model could be modified to diagnose on the basis of change in a patient's variables rather than by comparison with a norm. This would improve diagnosis among persons suffering from one disease where the diagnosis is being complicated by the existence of another disease.

The extension of the model to include the diagnosis of women would require only a change in the prior probability

to include a test for sex. Additionally, a statistical check of the indicators would be necessary to determine if a new data base including women would be necessary if women were to be tested.

Once a person has been found to have CHD, a system to monitor his progress under dieting and exercise control could be developed from the present model. This could allow a technician rather than a doctor to periodically check the patient's indicators.

The definitions used for positive tests throughout this study were based on current information. Both a statistical and medical investigation in this area to better define test results could greatly improve future models developed on the same principles.

A model to predict the cost of implementing and operating the proposed diagnostic model should be explored

APPENDIX A: SAMPLE DATA COLLECTION SHEET

Name

Date

RACE: CAU NEG MON

Sex_____ Height_____ Blood Pressure_____

Age_____ Weight_____ Blood Type_____

Family History: Any of the following diagnosed heart diseases (circle)

Father	Uncle		
Mother	Brother	Unknown	None
Aunt	Sister		

Any of the following died of heart disease (circle)

Father	Uncle		
Mother	Brother	Unknown	None
Aunt	Sister		

Cigarette smoking in excess of one year? Yes___ No___

If yes: less than 1/2 pack per day
one pack per day
more than one pack per day

History of Ischemic episodes:

Chest pain, undetermined origin
Typical angina
None

Resting EKG:

Normal
ST-T abnormalities
Pathologic Q waves
Other

Exercise EKG:

Neg
ST depression greater than 1 mm
ST depression greater than 2 mm
ST elevation

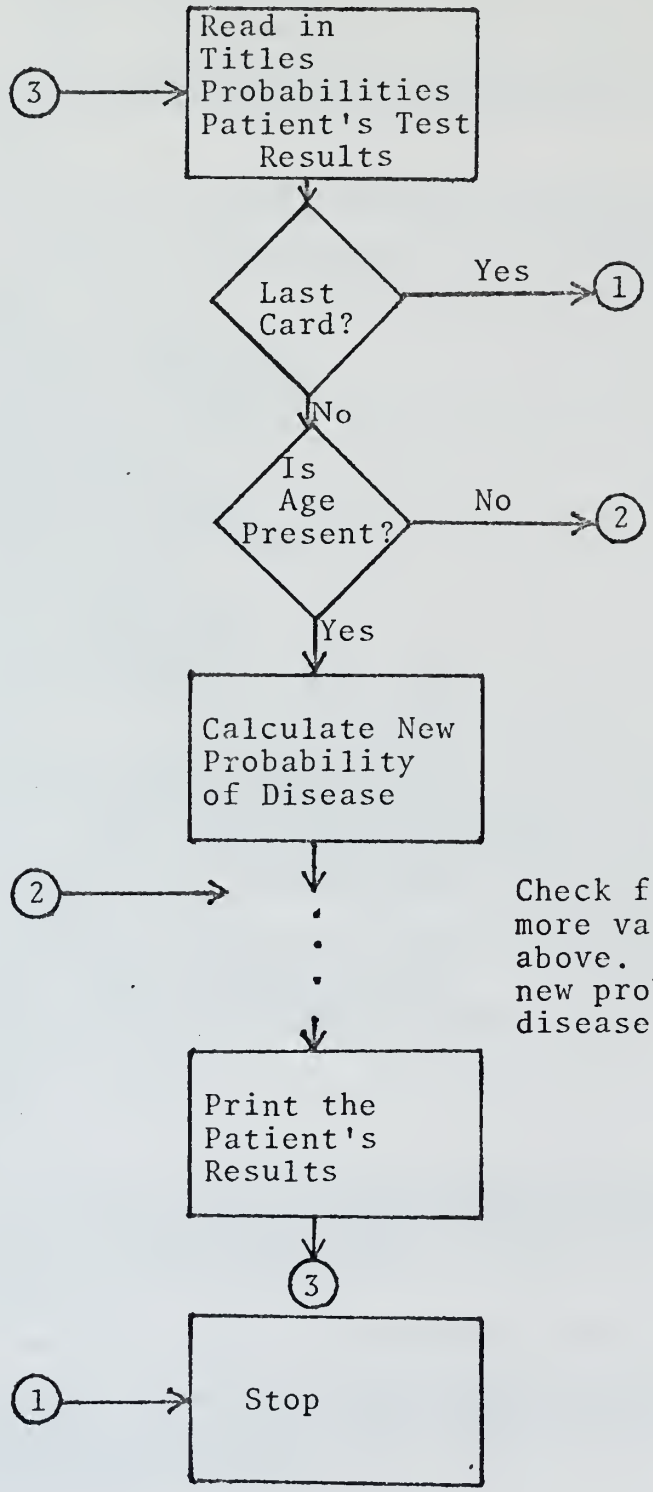
Triglycerides_____

Cholesterol_____

Max Heart Rate Attained during Exercise Test_____

APPENDIX B

BAYES' DIAGNOSTIC MODEL, FORTRAN FLOW CHART



APPENDIX C: BAYES' DIAGNOSTIC MODEL FORTRAN PROGRAM LISTING

```
    BAYE'S DIAGNOSTIC MODEL FOR CORONARY HEART DISEASE



    REAL*8B,C,D,E,F,G,S,T,U,V,BC,CD,DE,EF,FG,GS,ST,TU,UV,
   1W,XY,YZ,ZY,YX,ZF,ZW,ZM,ZB,ZH,ZO,ZG,ZN,ZC,ZD,VW,A,ZA

    READ IN TITLES FOR OUTPUT

    READ(5,99)A,B,C,D,E,F,G,S,T,U,V,W,BC,CD,DE,EF,FG,GS,
   1ST,TU,UV,VW,XY,YZ,ZY,YX,ZF,ZW,ZM,ZA,ZB,ZH,ZC,ZG,ZN,ZC,
   1ZD
 99 FORMAT(10A8)

    READ IN PROBABILITIES OF SYMPTOMS GIVEN NO CHD

    CODE FOR INPUT OF PROBABILITIES OF SYMPTOMS GIVEN
    NO CHD (AND CHD)
    1ST LETTER
       P-PROBABILITY OF
    2ND AND 3RD LETTER
       BD-DIASTOLIC PRESSURE
       BS-SYSTOLIC PRESSURE
       BT-BLOOD TYPE
       CI-CIGARETTE HABITS
       CH-CHOLESTEROL
       EE-EXERCISE EKG
       FN-FAMILY HISTORY
       HE-HISTORY OF ISCHEMIA
       RE-RESTING EKG
       RN- RACE NEGRO
       RC-RACE CAUCASIAN
       RM-RACE MONGOLIAN
       TY-TRIGLYCERIDE
    4TH LETTER IN A FOUR LETTER CODE
       D-CHD
       N-NO CHD
    4TH LETTER IN A FIVE LETTER CODE
       A-ABNORMAL WHEN PRECEDED BY RE, TY, CH, BS, OR BD
       A-ANGINA WHEN PRECEDED BY HE
       A-BLOOD TYPE A WHEN PRECEDED BY BT
       G-GREATER THAN 1MM DEPRESSION WHEN PRECEDED BY EE
       G-GREATER THAN 1 WHEN PRECEDED BY CI
       H-1/2 PACK
       N-NEGATIVE OR NONE
       O-BLOOD TYPE O, AB OR B WHEN PRECEDED BY BT
       O-1 PACK WHEN PRECEDED BY CI
       O-OTHER WHEN PRECEDED BY HE
       P-PAIN UND. ORIGIN WHEN PRECEDED BY HE
       P-POSITIVE
       Q-PATH. Q WAVES
    5TH LETTER
       D-CHD
       N-NO CHD

    READ(5,100)PRNN,PRCN,PRMN,PBTAN,PBTON,PFNPN,PFNNN,
   1PCIHN,PCION,PCIGN,PCINN,PHEPN,PHEAN,PHENN,PRENN,PREAN,
   1PREON,PREON,PEENN,PEEON,PEEGN,PTYNN,PTYAN,PCHAN,PCHNN,
   1PBSAN,PBSNN,PBDAN,PBDNN,PDA
100 FORMAT(8F10.6)

    READ IN PROBABILITIES OF SYMPTOMS GIVEN CHD

    READ(5,101)PRND,PRCD,PRMD,PBTAD,PBTOD,PFNPD,PFNND,
   1PCIHD,PCIOD,PCIGD,PCIND,PHEPD,PHEAD,PHEND,PREND,PREAD,
   1PREOD,PREOD,PEEND,PEEOD,PEEGD,PTYND,PTYAD,PCHAD,PCHND,
   1PBSAD,PBSND,PBDAD,PBDND
101 FORMAT(8F10.6)
    I=0.0
```

```
    READ IN PATIENT'S TEST RESULTS

    CODE FOR INPUT OF PATIENT'S TEST RESULTS
       INDIC- NOT USED COL.1
       R-RACE COL.6
            1-NEGRO
            2-CAUCASIAN
            3-MONGOLIAN
       Z- NOT USED COLS. 9-11
       BS-SYSTOLIC PRESSURE COLS. 14-16 NUMERICAL VALUE
       BD-DIASTOLIC PRESSURE COLS. 19-21 NUMERICAL VALUE
       BT-BLOOD TYPE COL. 24
            1-A
            2-O
            3-B
            4-AB
       FN-FAMILY HISTORY COL. 27
            1-POSITIVE
            2-NEGATIVE
       CI-CIGARETTE HABITS COL.30
            1-1/2 PACK
            2-ONE PACK
            3-GREATER THAN ONE
            4-NONE
       HIE-HISTORY OF ISCHEMIA COL. 33
            1-PAIN OF UNDETERMINED ORIGIN
            2-ANGINA
            3-NONE
       RE-RESTING EKG COL. 36
            1-NORMAL
            2-ABNORMAL S-T SEGMENT
            3-Q-WAVES
            4-OTHER
       EE-EXERCISE EKG COL. 39
            1-NORMAL
            2-LESS THAN 1/2 MM
            3-1/2 TO 1 MM DEPRESSION
            4-GREATER THAN 1MM DEPRESSION OR A ST ELEVATION
       TRY-TRIGLYCERIDE COL. 42
            1-NORMAL
            2-ABNORMAL
       CHL- CHOLESTEROL COLS. 45-47 NUMERICAL VALUE
       AGE-AGE COLS.50-51 NUMERICAL VALUE

 50 READ(5,103)INDIC,R,Z,BS,BD,BT,FN,CI,HIE,RE,EE,TRY,CHL,
     1AGE
103 FORMAT(A1,4X,F1.0,2X,F3.0,2X,F3.0,2X,F3.0,2X,F1.0,2X,
     1F1.0,2X,F1.0,2X,F1.0,2X,F1.0,2X,F1.0,2X,F1.0,2X,F3.0,
     12X,F2.0)

    ZERO COUNT OF MISSING TESTS

    MAGE=0
    MR=0
    MBT=0
    MFN=0
    MCI=0
    MHE=0
    MRE=0
    MEE=0
    MTY=0
    MCL=0
    MBS=0
    MBD=0
    J=0
    I=I+1

    CHECK FOR LAST CARD

    IF(R.EQ.9)GO TO 5000
```

```
      CHECK FOR AGE

      IF(AGE.EQ.0.0)GO TO 200

      MAKE FIRST CHECK FOR AGE GROUP, LESS THAN 34, 34 TO
      44, OVER 44
      AFTER DETERMINATION ASSIGN PRIOR PROBABILITY OF CHD

      IF(AGE.GT.34)GO TO 120
      PD=.01
      GO TO 190
  120 IF(AGE.GT.44)GO TO 130
      PD=.04
      GO TO 190
  130 PD=.07
  190 WRITE(6,191)I
  191 FORMAT('0',/,23X,'SUBJECT #',2X,I3)
      GO TO 201
  200 PD=PDA
      MAGE=MAGE+1

      CHECK TO DETERMINE RACE, RECALCULATE PROBABILITY OF
      CHD

  201 IF(R.EQ.0.0)GO TO 301
      IF(R.EQ.1.0)PDR=(PRCD*PD)/((PRCD*PD)+(PRCN*(1.0-PD)))
      IF(R.EQ.2.0)PDR=(PRND*PD)/((PRND*PD)+(PRNN*(1.0-PD)))
      IF(R.EQ.3.0)PDR=(PRMD*PD)/((PRMD*PD)+(PRMN*(1.0-PD)))
      GO TO 302
  301 PDR=PD
      MR=MR+1

      CHECK TO DETERMINE BLOOD TYPE, RECALCULATE PROBABILITY
      CHD

  302 IF(BT.EQ.0.0)GO TO 401
      IF(BT.EQ.1.0)PDBT=(PBTAD*PDR)/((PBTAD*PDR)+
     1(PBTAN*(1.0-PDR)))
      IF(BT.EQ.2.0)PDBT=(PBTOD*PDR)/((PBTOD*PDR)+
     1(PBTON*(1.0-PDR)))
      IF(BT.EQ.3.0)PDBT=(PBTOD*PDR)/((PBTOD*PDR)+
     1(PBTON*(1.0-PDR)))
      IF(BT.EQ.4.0)PDBT=(PBTOD*PDR)/((PBTOD*PDR)+
     1(PBTON*(1.0-PDR)))
      GO TO 402
  401 PDBT=PDR
      MBT=MBT+1

      CHECK TO DETERMINE FAMILY HISTORY, RECALCULATE PROB-
      ABILITY OF CHD

  402 IF(FN.EQ.0.0)GO TO 501
      IF(FN.EQ.1.0)PDFN=(PFNPD*PDBT)/((PFNPD*PDBT)+
     1(PFNPN*(1.0-PDBT)))
      IF(FN.EQ.2.0)PDFN=(PFNND*PDBT)/((PFNND*PDBT)+
     1(PFNNN*(1.0-PDBT)))
      GO TO 502
  501 PDFN=PDBT
      MFN=MFN+1

      CHECK TO DETERMINE CIGARETTE HABITS,RECALCULATE PROB-
      ABILITY OF CHD

  502 IF(CI.EQ.0.0)GO TO 601
      IF(CI.EQ.1.0)PDCI=(PCIHD*PDFN)/((PCIHD*PDFN)+
     1(PCIHN*(1.0-PDFN)))
      IF(CI.EQ.2.0)PDCI=(PCIOD*PDFN)/((PCIOD*PDFN)+
     1(PCION*(1.0-PDFN)))
      IF(CI.EQ.3.0)PDCI=(PCIGD*PDFN)/((PCIGD*PDFN)+
     1(PCIGN*(1.0-PDFN)))
      IF(CI.EQ.4.0)PDCI=(PCIND*PDFN)/((PCIND*PDFN)+
     1(PCINN*(1.0-PDFN)))
```

```
      GO TO 602
  601 PDCI=PDFN
      MCI=MCI+1

      CHECK TO DETERMINE HISTORY OF ISCHEMIC EPISODES,RECAL-
      CULATE PROBABILITY OF CHD

  602 IF(HIE.EQ.0.0)GO TO 701
      IF(HIE.EQ.1.0)PDHE=(PHEPD*PDCI)/((PHEPD*PDCI)+
     1(PHEPN*(1.0-PDCI)))
      IF(HIE.EQ.2.0)PDHE=(PHEAD*PDCI)/((PHEAD*PDCI)+
     1(PHEAN*(1.0-PDCI)))
      IF(HIE.EQ.3.0)PDHE=(PHEND*PDCI)/((PHEND*PDCI)+
     1(PHENN*(1.0-PDCI)))
      GO TO 702
  701 PDHE=PDCI
      MHE=MHE+1

      CHECK TO DETERMINE RESTING EKG RESULTS,RECALCULATE
      PROBABILITY OF CHD

  702 IF(RE.EQ.0.0)GO TO 801
      IF(RE.EQ.1.0)PDRE=(PREND*PDHE)/((PREND*PDHE)+
     1(PRENN*(1.0-PDHE)))
      IF(RE.EQ.2.0)PDRE=(PREAD*PDHE)/((PREAD*PDHE)+
     1(PREAN*(1.0-PDHE)))
      IF(RE.EQ.3.0)PDRE=(PREQD*PDHE)/((PREQD*PDHE)+
     1(PREQN*(1.0-PDHE)))
      IF(RE.EQ.4.0)PDRE=(PREOD*PDHE)/((PREOD*PDHE)+
     1(PREON*(1.0-PDHE)))
      GO TO 802
  801 PDRE=PDHE
      MRE=MRE+1

      CHECK TO DETERMINE EXERCISE EKG RESULTS, RECALCULATE
      PROBABILITY OF CHD

  802 IF(EE.EQ.0.0)GO TO 901
      IF(EE.EQ.1.0)PDEE=(PEEND*PDRE)/((PEEND*PDRE)+
     1(PEENN*(1.0-PDRE)))
      IF(EE.EQ.2.0)PDEE=(PEEOD*PDRE)/((PEEOD*PDRE)+
     1(PEEON*(1.0-PDRE)))
      IF(EE.GE.3.0)PDEE=(PEEGD*PDRE)/((PEEGD*PDRE)+
     1(PEEGN*(1.0-PDRE)))
      GO TO 902
  901 PDEE=PDRE
      MEE=MEE+1

      CHECK TO DETERMINE TRIGLYCERIDES RESULTS, RECALCULATE
      PROBABILITY OF CHD

  902 IF(TRY.EQ.0.0)GO TO 925
      IF(TRY.EQ.1.0)PDTY=(PTYND*PDEE)/((PTYND*PDEE)+
     1(PTYNN*(1.0-PDEE)))
      IF(TRY.GE.2.0)PDTY=(PTYAD*PDEE)/((PTYAD*PDEE)+
     1(PTYAN*(1.0-PDEE)))
      GO TO 926
  925 PDTY=PDEE
      MTY=MTY+1

      CHECK TO DETERMINE CHOLESTEROL RESULTS, RECALCULATE
      PROBABILITY OF CHD

  926 IF(CHL.EQ.0.0)GO TO 950
      IF(AGE.GT.29.0)GO TO 930
      IF(CHL.GT.240)PDCL=(PCHAD*PDTY)/((PCHAD*PDTY)+
     1(PCHAN*(1.0-PDTY)))
      IF(CHL.LE.240)PDCL=(PCHND*PDTY)/((PCHND*PDTY)+
     1(PCHNN*(1.0-PDTY)))
      GO TO 951
  930 IF(AGE.GT.39.0)GO TO 935
      IF(CHL.GT.270)PDCL=(PCHAD*PDTY)/((PCHAD*PDTY)+
```

```
     1(PCHAN*(1.0-PDTY)))
      IF(CHL.LE.270)PDCL=(PCHND*PDTY)/((PCHND*PDTY)+
     1(PCHNN*(1.0-PDTY)))
      GO TO 951
  935 IF(AGE.GT.49.0)GO TO 945
      IF(CHL.GT.310)PDCL=(PCHAD*PDTY)/((PCHAD*PDTY)+
     1(PCHAN*(1.0-PDTY)))
      IF(CHL.LE.310)PDCL=(PCHND*PDTY)/((PCHND*PDTY)+
     1(PCHNN*(1.0-PDTY)))
      GO TO 951
  945 IF(CHL.GT.330)PDCL=(PCHAD*PDTY)/((PCHAD*PDTY)+
     1(PCHAN*(1.0-PDTY)))
      IF(CHL.LE.330)PDCL=(PCHND*PDTY)/((PCHND*PDTY)+
     1(PCHNN*(1.0-PDTY)))
      GO TO 951
  950 PDCL=PDTY
      MCL=MCL+1

      CHECK TO DETERMINE SYSTOLIC BLOOD PRESSURE, RECALCU-
      LATE PROBABILITY OF CHD

  951 IF(BS.EQ.0.0)GO TO 975
      IF(BS.GE.140)PDBS=(PBSAD*PDCL)/((PBSAD*PDCL)+
     1(PBSAN*(1.0-PDCL)))
      IF(BS.LT.140)PDBS=(PBSND*PDCL)/((PBSND*PDCL)+
     1(PBSNN*(1.0-PDCL)))
      GO TO 976
  975 PDBS=PDCL
      MBS=MBS+1

      CHECK TO DETERMINE DIASTOLIC BLOOD PRESSURE, RECALCU-
      LATE PROBABILITY OF CHD

  976 IF(BD.EQ.0.0)GO TO 990
      IF(BD.GT.90)PDBD=(PBDAD*PDBS)/((PBDAD*PDBS)+
     1(PBDAN*(1.0-PDBS)))
      IF(BD.LE.90)PDBD=(PBDND*PDBS)/((PBDND*PDBS)+
     1(PBDNN*(1.0-PDBS)))
      GO TO 991
  990 PDBD=PDBS
      MBD=MBD+1

      FORMAT OF OUTPUT INSTRUCTIONS

  991 WRITE(6,986)I
  986 FORMAT('0', 3X,'THE FOLLOWING INFORMATION IS MISSING',
     1' ON SUBJECT #',2X,I3)
      IF(MAGE.EQ.0)GO TO 993
      WRITE(6,992)VW
  992 FORMAT(26X,A8)
      GO TO 994
  993 J=J+1
  994 IF(MR.EQ.0)GO TO 996
      WRITE(6,992)A
      GO TO 997
  996 J=J+1
  997 IF(MBT.EQ.0)GO TO 998
      WRITE(6,995)B,C
  995 FORMAT(26X,A8,A8)
      GO TO 999
  998 J=J+1
  999 IF(MFN.EQ.0)GO TO 1000
      WRITE(6,995)D,E
      GO TO 1001
 1000 J=J+1
 1001 IF(MCI.EQ.0)GO TO 1002
      WRITE(6,995)F,G
      GO TO 1003
 1002 J=J+1
 1003 IF(MHE.EQ.0)GO TO 1004
      WRITE(6,995)S,T
      GO TO 1005
```

```
1004 J=J+1
1005 IF(MRE.EQ.0)GO TO 1006
     WRITE(6,995)U,V
     GO TO 1007
1006 J=J+1
1007 IF(MEE.EQ.0)GO TO 1008
     WRITE(6,995)W,BC
     GO TO 1009
1008 J=J+1
1009 IF(MTY.EQ.0)GO TO 1010
     WRITE(6,995)CD,DE
     GO TO 1011
1010 J=J+1
1011 IF(MCL.EQ.0)GO TO 1012
     WRITE(6,995)EF,FG
     GO TO 1013
1012 J=J+1
1013 IF(MBS.EQ.0)GO TO 1014
     WRITE(6,995)GS,ST
     GO TO 1015
1014 J=J+1
1015 IF(MBD.EQ.0)GO TO 1016
     WRITE(6,995)TU,UV
     GO TO 1017
1016 J=J+1
1017 IF(J.LT.12)GO TO 1020
     WRITE(6,1018)
1018 FORMAT('0',28X,'NONE')
1020 WRITE(6,1021)I,PDBD
1021 FORMAT('0',4X,'SUBJECT #',1X,I3,1X,'HAS PROBABILITY',
    12X,F8.6,2X,'OF CORONARY'/' ',5X,'HEART DISEASE GIVEN',
    1' RESULTS IN THE FOLLOWING TESTS')
     IF(MAGE.EQ.1)GO TO 1024
     WRITE(6,1022)VW,AGE
1022 FORMAT(18X,A8,13X,F3.0)
1024 IF(MR.EQ.1)GO TO 1026
     IF(R.EQ.1.0)WRITE(6,1025)A,ZW
     IF(R.EQ.2.0)WRITE(6,1025)A,ZF
     IF(R.EQ.3.0)WRITE(6,1025)A,ZM
1025 FORMAT(18X,A8,13X,A8)
1026 IF(MBT.EQ.1)GO TO 1029
     IF(BT.EQ.1.0)WRITE(6,1028)B,C,ZA
     IF(BT.GT.1.0)WRITE(6,1028)B,C,ZB
1028 FORMAT(18X,A8,A8,5X,A8)
1029 IF(MFN.EQ.1)GO TO 1030
     IF(FN.EQ.1.0)WRITE(6,1028)D,E,ZY
     IF(FN.EQ.2.0)WRITE(6,1028)D,E,YX
1030 IF(MCI.EQ.1)GO TO 1031
     IF(CI.EQ.1.0)WRITE(6,1028)F,G,ZH
     IF(CI.EQ.2.0)WRITE(6,1028)F,G,ZO
     IF(CI.EQ.3.0)WRITE(6,1028)F,G,ZG
     IF(CI.EQ.4.0)WRITE(6,1028)F,G,ZN
1031 IF(MHE.EQ.1)GO TO 1032
     IF(HIE.EQ.1.0)WRITE(6,1028)S,T,ZC
     IF(HIE.EQ.2.0)WRITE(6,1028)S,T,ZD
     IF(HIE.EQ.3.0)WRITE(6,1028)S,T,ZN
1032 IF(MRE.EQ.1)GO TO 1033
     IF(RE.EQ.1.0)WRITE(6,1028)U,V,XY
     IF(RE.GE.2.0)WRITE(6,1028)U,V,YZ
1033 IF(MEE.EQ.1)GO TO 1034
     IF(EE.EQ.1.0)WRITE(6,1028)W,BC,XY
     IF(EE.GE.2.0)WRITE(6,1028)W,BC,YZ
1034 IF(MTY.EQ.1)GO TO 1035
     IF(TRY.EQ.1.0)WRITE(6,1028)CD,DE,XY
     IF(TRY.GE.2.0)WRITE(6,1028)CD,DE,YZ
1035 IF(MCL.EQ.1)GO TO 1037
     WRITE(6,1036)EF,FG,CHL
1036 FORMAT(18X,A8,A8,5X,F4.0)
1037 IF(MBS.EQ.1)GO TO 1038
     WRITE(6,1036)GS,ST,BS
1038 IF(MBD.EQ.1)GO TO 50
     WRITE(6,1036)TU,UV,BD
```

```
      GO TO 50
5000  STOP
      END
```

APPENDIX D

SAMPLE OUTPUT

SUBJECT # 7

THE FOLLOWING INFORMATION IS MISSING ON SUBJECT # 7

RACE
BLOOD TYPE
RESTING EKG
TRIGLYCERIDE
CHOLESTEROL

SUBJECT # 7 HAS PROBABILITY 0.979999 OF CORONARY HEART DISEASE GIVEN RESULTS IN THE FOLLOWING TESTS

AGE	44.
FAMILY HISTORY	NEGATIVE
SMOKING HABITS	NONE
ISCHEMIA HISTORY	ANGINA
EXERCISE EKG	ABNORMAL
SYSTOLIC PRESS.	150.
DIASTOLIC PRESS.	100.

SUBJECT # 8

THE FOLLOWING INFORMATION IS MISSING ON SUBJECT # 8

NONE

SUBJECT # 8 HAS PROBABILITY 0.999946 OF CORONARY HEART DISEASE GIVEN RESULTS IN THE FOLLOWING TESTS

AGE	28.
RACE	WHITE
BLOOD TYPE	A
FAMILY HISTORY	NEGATIVE
SMOKING HABITS	OVER ONE
ISCHEMIA HISTORY	ANGINA
RESTING EKG	ABNORMAL
EXERCISE EKG	ABNORMAL
TRIGLYCERIDE	ABNORMAL
CHOLESTEROL	315.
SYSTOLIC PRESS.	118.
DIASTOLIC PRESS.	76.

LIST OF REFERENCES

1. Blackburn, H., Measurement in Exercise Electrocardiography, Thomas, 1969.

2. Brodman, K., "Diagnostic Decisions by Machine," IRE Transactions on Medical Electronics, July 1960.

3. Chan, L. S. and Dunn, O. J., "The Treatment of Missing Values in Discriminant Analysis - 1. The Sampling Experiment," Journal of the American Statistical Association, v. 67, n. 338, June 1972.

4. Cohn, P. F., and others, "A Quantitative Clinical Index for the Diagnosis of Symptomatic Coronary-Artery Disease," The New England Journal of Medicine, v. 286, n. 17, 27 April 1972.

5. Dunn, O. J., "Some Expected Values for Probabilities of Correct Classification in Discriminant Analysis," Technometrics, v. 13, n. 2, May 1971.

6. Forsyth, J. W., A Manual of Exercise Testing, Medford Clinic, Medford, Oregon, January 1972.

7. Gessaman, M. P. and Gessaman, P. H., "A Comparison of Some Multivariate Discrimination Procedures," Journal of the American Statistical Association, v. 67, n. 338, June 1972.

8. Gustafson, D. H., and others, "Initial Evaluation of a Subjective Bayesian Diagnostic System," Health Services Research, Fall 1971.

9. Kingsbury, K. J., "Relation of ABO Blood-Groups to Atherosclerosis," The Lancet, 30 January 1971.

10. Lusted, L. B., Introduction to Medical Decision Making, Thomas, 1959.

11. Nie, N., Beut, D. H., and Hull, C. H., Statistical Package for the Social Sciences, McGraw-Hill, 1970.

12. Page, I. H., and others, "Prediction of Coronary Heart Disease Based on Clinical Suspicion, Age, Total Cholesterol, and Triglyceride," Circulation, v. XLII, October 1970.

13. Rao, C. R., "Recent Advances in Discriminatory Analysis," Indian Society of Agricultural Statistics Journal, v. 21, 1969.

14. Stamler, J., and others, "Primary Prevention of the Atherosclerotic Diseases," Circulation, v. XLII, December 1970.

15. Stanford University Department of Statistics Report 10, Asymptotic Evaluation of the Probabilities of Mesclassification by Linear Discriminant Functions, by T. W. Anderson, 28 September 1972.

16. "The Heart Attack Epidemic," Life Extension, v. 3, n. 7, 1972.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Chief of Naval Personnel Pers 11b Department of the Navy Washington, D. C. 20370	1
4.	Naval Postgraduate School Department of Operations Research and Administrative Sciences Monterey, California 93940	1
5.	Asst Professor Marlin U. Thomas, Code 55 To Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
6.	Asst Professor Edward A. Brill, Code 55 Zg Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
7.	Asst Professor William C. Giauque, Code 55 Gi Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
8.	The Surgeon General Department of the Army Washington, D. C. 20314	1
9.	Commanding Officer Silas B. Hays Army Hospital Fort Ord, California 93941	1
10.	Chief of Cardiology Walter Reed General Hospital Washington, D. C. 20012	1

11.	Major James A. Fischer, M.D. Cardiology Clinic Silas B. Hays Army Hospital Fort Ord, California 93941	1
12.	Captain William R. Condos, Jr. Department of Mathematics U.S. Military Academy West Point, New York 10996	2
13.	Captain Everett W. Knox 190 Glascoe Avenue Staten Island, New York 10314	2

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author)

Naval Postgraduate School
Monterey, California 93940

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

REPORT TITLE

A Bayesian Approach to Assist in the Diagnosis of Coronary Heart Disease

DESCRIPTIVE NOTES (Type of report and inclusive dates)

Master's Thesis; March 1973

AUTHOR(S) (First name, middle initial, last name)

William Randolph Condos, Jr. and Everett William Knox

REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1973	47	16

a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

The objectives of this thesis were to design a method for evaluation of the diagnostic potential of available indicators of coronary heart disease (CHD) and to present a systematic, quantitative procedure for aiding in its diagnosis. A sample space of patients was divided into two mutually exclusive groups, those with angiographic evidence of CHD, and those with no CHD. Active duty or retired military men between the ages of 30 and 67 years constituted the sample space. Tests and risk factors were available in the medical literature that a doctor could view as an indicator or contraindicator of CHD. A vector of these possible indicators was established and the diseased group was compared to the non-diseased group in an effort to evaluate the diagnostic potential of the indicators. This was done by discriminant analysis in conjunction with a Bayesian method of weighting the importance of test results. The important indicators were then used to formulate a model for diagnosing CHD based on a Bayes' decision technique.

DD FORM 1 NOV 68 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

4. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Bayes						
Quantitative Analysis						
Discriminant Analysis						
Coronary Heart Disease						
Medical Decision Making						

Thesis
C669
c.1

143274

Condos
A Bayesian approach to assist in the diagnosis of coronary heart disease.